

⑨ 日本国特許庁 (JP)

⑩ 特許出願公開

⑪ 公開特許公報 (A)

昭58—4198

⑫ Int. Cl.³
G 10 L 1/00

識別記号

厅内整理番号
7350-5D

⑬ 公開 昭和58年(1983)1月11日

発明の数 1
審査請求 未請求

(全 4 頁)

④ 音声認識装置における標準パターン登録方式

② 特願 昭56—102035

② 出願 昭56(1981)6月30日

② 発明者 高橋次男

小田原市国府津2880番地株式会
社日立製作所小田原工場内

② 発明者 栗野清道

小田原市国府津2880番地株式会

社日立製作所小田原工場内
② 発明者 高木賢一

小田原市国府津2880番地株式会

社日立製作所小田原工場内

② 出願人 株式会社日立製作所

東京都千代田区丸の内1丁目5
番1号

② 代理人 弁理士 鈴木誠

明細書

発明の名称

音声認識装置における標準パターン登録方式
特許請求の範囲

1. 話者の発声した音声（被認識音声）のパターンと、予め登録しておいた標準パターンとの類似度を調べることにより、被認識音声を認識する音声認識装置において、標準パターンの登録にあたって、同一話者により同一カテゴリの音声を2回以上続けて発声させ、当該音声認識装置自体で、各回の音声のパターン相互の類似度を調べ、所定以上の類似度が得られたいずれかの回の音声のパターンを該当カテゴリの標準パターンとして登録することを特徴とする標準パターン登録方式。

発明の詳細な説明

本発明は、特定の話者を対象とした音声認識装置に関し、特に、話者の発声した音声を認識するために、その音声のパターンと比較される標準パターンを登録する方式に関する。

音声認識装置は、コンピュータの有望な入力装置として脚光を浴びているが、その中で実用段階に最も近いのは、現在のところ、特定の話者を対象にした音声認識装置である。このような特定話者対象の音声認識装置は、普通、認識対象の話者の発声した音声を周波数分析して音声パターンを得、それを標準パターンとして予め登録しておき、その後は、この標準パターンを用いて話者発声の音声のパターンとの類似度判定を行なうことにより、音声を認識するように構成されている。

ところで、かかる従来の音声認識装置においては、予め決めた話などを記者によって発声させ、その音声パターンをそのまま標準パターンとして登録している。そして、適切な標準パターンが登録されたか否かのチェックは、実際に認識動作を行なわせ、安定した認識が行なわれるか否かによって判断している。しかし、これでは信頼性を十分に確保する上で問題があつた。

そこで、信頼性を向上するために、同一カテゴリの音声に対する標準パターンを複数組登録する

ことも考えられている。しかしながら、これでは標準パターンのファイルが大形になってしまい、また、認識の際に実際には使用する音声のカテゴリ数の倍以上の標準パターンに対して類似判定を実行しなければならないため、認識速度の低下やコスト上昇を招く。その上、必ずしも信頼性の改善をそれ程期待できない。何故なら、同一カテゴリについていくつ標準パターンを登録しても、その中に標準パターンとして適正なパターンが必ず含まれるという保証は無いからである。

したがつて本発明の目的は、音声認識装置の信頼性を改善でき、しかもし記のような問題を伴なわない、標準パターンの登録方式を提供することにある。

しかして本発明の標準パターンの登録方式の特徴は、話者に同一カテゴリの音声を2回以上続けて発声させ、音声認識装置自体で各回の音声パターン同士の類似判定を行ない、所定以上の類似度が得られたいずれかの回の音声パターンを標準パターンとして登録する点にある。

以下、本発明を一実施例について詳細に説明する。

本発明により標準パターン登録を実施するよう構成した音声認識装置の一例を第1図に示し、説明する。

1はマイクロフォンであり、話者によつて発声された音声はこのマイクロフォン1により電気信号に変換され、装置内に取り込まれる。音声信号は、送信増幅器2を通じてロチネルの帯域フィルタ部3に入力される。帯域フィルタ部3のチャネル数は通常16～20チャネルであり、帯域フィルタ部3の各チャネルのフィルタ3₁～3_nには約200Hz～6000Hzの周波数域を適切な帯域に分割して周波数割付けされている。帯域フィルタ部3の各チャネルの出力は、帯域フィルタ部4のフィルタ4₁～4_nにチャネル対応で入力される。かくして、帯域フィルタ部3で音声信号から分離抽出した各周波成分の複数のエンベロープが得られ、アナログスイッチ5に送られる。帯域フィルタ部4の各チャネルのフィルタ4₁～4_nは、時定数50ms

程度の低域フィルタが用いられる。

アナログスイッチ5は、制御部9の制御下で、帯域フィルタ部4の各チャネル出力を約20ms毎にA/Dコンバータ6へ時分割で送る。A/Dコンバータ6は、アナログスイッチ5より入力される信号を8ないし16ビットのデジタル信号に変換する(つまり、2⁸ないし2¹⁶レベルにデジタル化する)。そして、チャネル1～チャネルnのA/Dコンバータ出力の組が、約20ms毎に1フレーム分のスペクトル情報をバッファメモリ7に一時的に記憶される。

演算部8は、バッファメモリ7から与えられるスペクトル情報を正規化する。この正規化は、入力音声の強さのバラつきを補償し、強さに影響されないスペクトルを得るために行なうもので、例えば次式(I)の演算によりなされる。

$$\frac{P_i}{\sum_{i=1}^n P_i} \times C \quad \dots \text{式(I)}$$

ここで、P_iはチャネルiのスペクトル情報のパワー、nは全チャネル数、Cは定数である。つま

り、バッファメモリ7から与えられるスペクトル情報の各フレームのトータルパワーが、定数Cに等しくなるように正規化することになる。

演算部8はまた、制御部9の制御の下に音声区間の切出しも行なう。具体的には、正規化前のスペクトル情報のフレーム毎のトータルパワーを、予め設定した閾値と比較し、この閾値以上のフレームを音声信号のフレームと判定し、閾値未満のフレームをノイズ信号のフレームと判定する。そして、音声信号のフレームが5フレーム連続すると、その最初のフレームより音声区間とみなし、その後、ノイズ信号のフレームが10フレームに連続すると、その最初のフレームを音声区間の終点とみなす。演算部8はこのようにして音声区間を検出し、この音声区間内のスペクトル情報(正規化後)を入力音声のパターンとして切り出す。

演算部8より出力される音声パターンは、通常の認識動作のときはマクシング部10にのみ入力されるが、標準パターンの登録時はランダムアクセスファイル(RAMと略記する)10にも選択的に

送られ格納される。このRAM10としては、磁気ディスクやフロッピーディスク等が使用される。

DPマッチング部IIは、制御部9の制御の下に、演算部8より与えられる入力音声のパターンと、RAM10より読み出される標準パターンとの間で、いわゆる動的計画法(Dynamic Programming; DP)に基づくパターンマッチングを行なう。特定の話者にあつては、音声の周波数スペクトルが安定している前面、発声速度のバラツキはかなり大きく、これを収めてパターン間の整合距離を求めるのにDPマッチング手法が有効である。DPマッチングは音声認識分野で周知があるので、これ以上の説明は略す。

さて、入力並声パターンと標準パターンとの間の整合距離がDPマッチング部IIで求められ、判定部12に入力される。通常の認識動作においては、判定部12は制御部9の制御下で、入力音声パターンと整合距離の最も小さな標準パターンを第1判定候補として、その次に小さな標準パターンを第2判定候補として選び、両候補の入力音声

パターンとの整合距離の差が規定値以上であれば、第1判定候補の標準パターンのコードを入力音声の認識結果として出力する。そうでなければ、入力音声を認識できないとして、判定部12はリジエクト出力を送出する。標準パターンの登録時ににおける判定部12の動作については、後述する。

今までの説明で、入力音声の認識時の動作は明らかであらうから、次に、標準パターンの登録に限定して説明する。

当該実施例においては、あるカテゴリの音声の標準パターンを登録するには、話者がその音声を2回続けて発声する。まず1回目に発声された音声のパターンが前述のようにして求められ、それがRAM10の該当カテゴリの領域に標準パターンとして仮登録される。次に、2回目に発声された音声のパターンと、標準パターンとしてRAM10に仮登録されている1回目の音声パターンとの整合距離がDPマッチング部IIで求められる。この整合距離が予め決められた判定閾値以下であるか否かの判定が、判定部12で行なわれる。整合距離

が判定閾値以下であれば、仮登録の音声パターンが有効な標準パターンとして本登録されるが、判定閾値を超えると判定されると、仮登録パターンを無効にする。具体的には、無効の判定信号が判定部12より出ると、制御部9の制御下でRAM10の該当エリアに無効コードが書き込まれたり、あるいは無効を表示するフラグが用意されているならば、その無効フラグがセットされる。

標準パターンの登録処理の流れ図を第2図に示す。この例は、RAM10の各カテゴリのエリアに無効フラグを用意してある場合である。

すなわち、カテゴリAの標準パターンを登録する場合、1回目に発声された音声のパターンA₁がRAM10に登録(この時点では仮登録)され、これと2回目に発声された音声のパターンA₂との整合距離S(A₁-A₂)が算出される。そして、この整合距離S(A₁-A₂)と判定閾値S₁とが比較され、S(A₁-A₂) ≤ S₁なら、パターンA₁がカテゴリAの有効な標準パターンとして判定されてRAM10に本登録され、カテゴリAの登録は終了

する。S(A₁-A₂) > S₁なら、RAM10のカテゴリAのエリアの無効フラグがセトされ、仮登録されたパターンA₁は無効となり、このカテゴリAの標準パターンの登録を始めからやり直す。なお、パターンA₁、A₂のいずれについても、人力直後に発声長(音声区間の長さ)のチェックが行なわれ、規定範囲から外れる場合は、そのパターンを再入力させる。この発声長のチェックは、演算部8において行なわれる。

尚、第1回の制御部9、演算部8、DPマッチング部IIは、純然たるハードウェア回路によって構成してもよいし、マイクロコンピュータを使用してもよい。たゞし、マイクロコンピュータの利用が有利な場合が多い。

前記実施例では、同一カテゴリの音声を2回続けて話者に発声させ、1回目と2回目のパターンの整合距離が所定閾値以下であれば、1回目のパターンを標準パターンとした。しかし、同一カテゴリの音声の連続発声回数は2回に限らず、3回以上とすることも許される。例えば、3回連続して

発声させ、1回目と2回目のパターン間の整合距離、2回目と3回目のパターン間の整合距離（またさらに、1回目と3回目のパターン間の整合距離）について、それぞれ前述のような判定を行ない、整合距離が所定値以下でかつ最も小さな値となつた組のいづれかの間のパターンを標準パターンとして登録する等のやり方も可能である。

本発明は以上に述べたように、音声認識装置自体で話者により連続して発声される音声のパターン同士の類似度を調べ、所定以上の類似度が得られたパターンを標準パターンとして登録する。間をおかずして続けて発声した場合、確めて類似した発声つまり音声のパターンを期待できるから、各回の音声パターンの類似度判定により、周囲騒音などの影響による音声区間の切り出し不良などを確實にチェックできる。また、2回またはそれ以上続けて発声した音声のパターンの中、類似したパターンを選択するから、1回だけの発声による場合よりも、より標準的なパターンを複数できる可能性が増す。また、標準パターンの候補に際して、

人間の判断に頼らないから、人为的なエラーにより不適切な標準パターンが登録されるというようなことも無くなる。

このように、本発明によれば、不適切な標準パターンの登録を排除できるため、音声認識装置の信頼度を高めることができる。また本発明によれば、標準パターンの登録作業が容易かつ迅速になるという効果もある。

図面の簡単な説明

第1図は本発明の一実施例である音声認識装置のプロック図、第2図は同上実施例における標準パターンの登録処理の流れの一例を示す図である。

1…マイクロフォン、2…接続増幅器、3…帯域フィルタ部、4…低域フィルタ部、5…アナログスイッチ、6…A/Dコンバータ、7…パッファメモリ、8…演算部、9…制御部、10…ランダムアクセスファイル（RAM）、11…D/Pマッピング部、12…判定部。

代理人 井丸士 鈴木 誠

第2図

第1図

